UNIVERSIDAD AUTÓNOMA DE MADRID
ESCUELA POLITÉCNICA SUPERIOR

Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación

TRABAJO FIN DE GRADO

Detección de ritmo cardiaco mediante análisis de secuencias de video en modalidad sin color.

Claudia Fernández Refoyo
Tutor: José María Martínez

Junio 2018

Detección de ritmo cardiaco mediante análisis de secuecias de video en modalidad sin color.

Claudia Fernández Refoyo
Tutor: José María Martínez

Video Processing and Understanding Lab
Departamento de Tecnología Electrónica y de las Comunicaciones
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Junio 2018

Trabajo parcialmente financiado por el Ministerio de Economía y Competitividad del Gobierno de España bajo el proyecto TEC2014-53176-R (HAVideo) (2015-2017)

Resumen

El objetivo de este trabajo fin de grado consiste en mejorar un algoritmo que permite detectar el ritmo cardiaco a partir de análisis de secuencias de vídeo basadas en cambios movimiento. Este algoritmo ha sido adaptado también para poder aplicarse en secuencias de profunidad. Se ha decidido utilizar el algoritmo en este tipo de secuencias porque no se ven influenciadas por las condiciones de iluminación y además garantizan la privacidad del individuo.
Se ha desarrollado por lo tanto un algoritmo en el que se obtiene el pulso de forma manual. Una vez que se ha validado el algoritmo de forma manual en secuencias de profundidad obteniéndose unos resultados bastante precisos, se ha procedido a añadir nuevas implementaciones para realizar el algoritmo de forma automática. Estas implementaciones se han realizado en la primera etapa del algoritmo: la detección de la región de interés. En este tipo de método se ha considerado también la distancia a la que se encuentra el individuo de la cámara.
Finalmente, se ha realizado una comparativa entre las dos técnicas desarrolladas concluyéndose que aunque se obtienen resultados más exactos con la forma manual, resulta mejor utilizar el algoritmo automático ya que no requiere de la interacción del usuario para reconocer la región de interés.

Palabras clave

Ritmo cardiaco, pulso, región de interés, puntos de interés, análisis de secuencias de video, cambios de movimiento, *datset*, imágenes de profundidad, detección manual, detección automática, Kinect, Kinect 2.0

Abstract

The objective of this Final Degree Thesis is to improve an algorithm that calculate the heart rate. This heart rate is detected by using depth images based on movement changes. The algorithm has been adapted to be applied on depth images. It has been decided to apply this algorithm on depth images because they can't be influence by the lighting conditions and they could ensure the individual privacy.
It has been implemented in the first instance an algorithm that could detect the heart rate manually. Once this algorithm has been valued on depth images by obtaining satisfactory results, it has been decided to add new implementations. This implementations have been incorporated on the first stage of the algorithm: the detection of the area of interest. This region of interest is detected automatically and considering the distance between the subject and the camera.
Lastly, it has been realized a comparative among the two methods. It is concluded that even thought the results obtained manually are more precises, it is better to use the algorithm automatically since it doesn't require the interaction of the person to indentify the area of interest.

Keywords

heart rate, heartbeat , area of interest, interesting points, video-sequences analysis, movement changes, dataset, depth images, manual detection, automatically detection, Kinect, Kinect 2.0

Agradecimientos

A la primera persona que se lo quiero agradecer es a mi tutor, Chema, por haberme dado la posibilidad de realizar este trabajo y por la ayuda y el conocimiento que me he aportado para desarrollarlo.
A mis padres, agradecerles todo el apoyo y cariño que me han dado durante estos maravillosos años de carrera. Agradecerles el esfuerzo y el sacrificio que han realizado para que haya podido tener siempre la oportunidad de estudiar y poder llegar hasta donde he llegado tras un largo camino de esfuerzo y trabajo.
En especial a mi hermano, por haber intentado sacarme siempre una sonrisa en los peores momentos que he atravesado durante este largo proceso.
A mi prima Belén y a mis abuelos por haberme apoyado siempre desde mi comienzo en la carrera y animarme en todo momento.
Al VPU y a la gente que me ha acompañado durante este largo recorrido, en especial a Ana, Raul, y Ariana.
A mis compañeros de clase durante todos estos años, en especial a Miguel, Emilio, Sergio, Alvar, Patricia y Julia. Agradecerles por todos los momentos que hemos compartido y que espero que sean muchos más.
A mis compañeros de trabajo, por haberse preocupado y haberme dado el apoyo y la comprensión durante los últimos meses tan duros que he atravesado. También agradecerles todos los conocimientos que me han aportado y que me han ayudado a crecer y a adquirir una gran visión para realizar este trabajo.
Y por último, a Jon, por haber sido la persona que más cosas me ha enseñado durante este último y duro año para mí y que desde luego nunca olvidaré.

Claudia Fernández Refoyo

Junio 2018

Índice general

Índice de figuras

Índice de tablas

Capítulo 1

Introducción

1.1. Motivación

Conocer el ritmo cardiaco puede ser vital para detectar problemas de salud, además de ayudar a evaluar la eficiencia física en cualquier deporte. Es una excelente forma de valorar si nuestro organismo está funcionando correctamente.
Hoy en día, muchos de los sistemas que se emplean para determinar el ritmo cardiaco necesitan aparataje o utilizan un sensor que lleva a cabo la medida del pulso cardiaco. Esto hace que de alguna manera el paciente tenga que verse involucrado en el proceso y tenga que entrar en contacto con estos aparatos. Por lo tanto, encontrar y desarrollar una metodología que permita obtener el valor del pulso sin utilizar aparatos *wearables* para procesarlo puede ser muy útil y novedoso. Existen técnicas que no requieren ni de aparatos ni de sensores y que se basan en análisis de video. Este tipo de sistemas pueden ser muy útiles entre otras cosas para la vigilancia de ancianos o de recién nacidos, pues su piel sensible podría ser dañada debido a la frecuente colocación de aparatajes.
Podemos clasificar los sistemas de medición de pulso basados en video en dos tipos: fundamentados en cambios de color y en cambios de movimiento. Los primeros detectan el ritmo cardiaco a partir de secuencias de video en color realizando un seguimiento de las variaciones de color [24]. Este tipo de métodos se basan en que es posible obtener una señal de pulso amplificando la variación periódica del color que presentan las secuencias que utilizan. Sin embargo, tienen como mayor desventaja que no protegen la privacidad del individuo bajo estudio, además de no funcionar eficientemente en condiciones de baja o nula iluminación.
Por ello, en este trabajo se propone utilizar técnicas basadas en cambios de movimiento. Se propone un algoritmo que detectará el pulso a partir de secuencias de

video realizando el análisis de los micro-movimientos inducidos por el bombeo de la sangre. Este algoritmo empleará como método de análisis el seguimiento de puntos característicos que serán extraídos de secuencias de profundidad. Estas secuencias tienen la capacidad de monitorizar situaciones incluso en condiciones de baja o nula iluminación.

1.2. Objetivos

Los objetivos propuestos para llevar a cabo este proyecto son los siguientes:

- Estudio del estado del arte.
- Desarrollo de un procedimiento de valoración de resultados.
- Implementación del algoritmo propuesto.
- Evaluación de un sistema que sea capaz de detectar la región de interés de secuencias de video en modalidad sin color.
- Desarrollo de dos técnicas para detectar la región de interés: de forma automática y de forma manual.
- Implementación en la técnica de detección de región de interés automática de un método que detecta dicha región teniendo en cuenta la distancia a la que se encuentra el sujeto de la cámara.

1.3. Organización de la memoria

La memoria consta de los siguientes capítulos:

- Capítulo 1. Motivación, propósitos y organización de la memoria.
- Capítulo 2. Estado del arte en la obtención del pulso cardiaco a través de secuencias de profundidad.
- Capítulo 3. Explicación del algoritmo propuesto y de los métodos implementados para detectar la región de interés.
- Capítulo 4. Explicación del *dataset* desarrollado. Explicación de las pruebas que han sido realizadas y de los resultados obtenidos en estas.
- Capítulo 5. Conclusiones y trabajo futuro.
- Bibliografía.

Capítulo 2

Estado del arte

2.1. Introducción

Para llevar a cabo este trabajo se han investigado tres puntos fundamentales.
En primer lugar, es importante saber cómo funciona nuestro corazón y como éste genera el pulso, pues de esta manera podemos comprender la trascendencia que puede tener utilizar este algoritmo.
En segundo lugar, se ha indagado acerca de las técnicas existentes para detectar el ritmo cardiaco mediante el análisis de video. Las primeras investigaciones que se basaban en video para detectar el pulso analizaban los cambios de color presentes en este. No obstante, estos cambios de color no eran perceptibles por el ojo humano por lo que no se podía discurrir que dichos métodos fueran del todo concluyentes. Fue entonces, cuando se publicó un trabajo que demostraba que empleando técnicas de amplificación de vídeo podían percibirse los cambios de color y también los cambios de movimiento [22]. Este trabajo fue el desencadenante de la elaboración de nuevas referencias basadas en color [12, 17, 20, 23, 3, 4, 8] y seguidamente en el movimiento [2]. Esto hizo que surgieran referencias para detectar el pulso cardiaco a partir de video basadas en movimiento [2]. Estos métodos analizan los micro-movimientos que se producen en la región facial causados por el flujo sanguíneo. Este trabajo se centrará en las técnicas basadas en movimiento. Estos procedimientos realizan el seguimiento de los puntos característicos extraídos de una región de interés. Por lo tanto, esta metodología nos permite poder usar secuencias de imágenes sin color ya que se basan en el movimiento y no en los cambios de color del video.
Y por último, se ha investigado también acerca del sensor Kinect v2 y de las imágenes de profundidad que dicho dispositivo proporciona y que hemos empleado para realizar este trabajo.

2.2. Funcionamiento del Corazón

El sistema vascular es el encargado de hacer llegar la sangre a cada una de las células del cuerpo humano. Después, la sangre regresa al corazón para ser revitalizada y se vuelve a repetir este proceso continuamente. Es decir, el corazón es el músculo que se encarga de bombear la sangre alrededor del cuerpo. Por el lado izquierdo del corazón llega la sangre rica en oxígeno procedente de los pulmones que posteriormente será suministrada al resto del cuerpo. Mientras que la sangre con insuficiente oxígeno regresa por el lado derecho del corazón donde se bombea de regreso a los pulmones para recibir oxígeno de nuevo. En la figura 2.1 se puede observar el procedimiento explicado anteriormente.

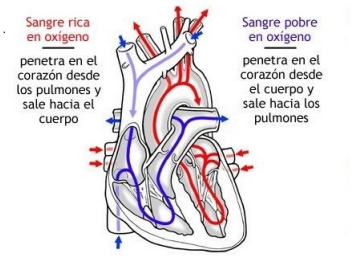


Figura 2.1: Ciclo que realiza la sangre en el corazón

De esta forma, el recorrido sanguíneo se produce con la sístole (contracción del corazón) y la diástole (dilatación del músculo cardíaco). El corazón dispone además de una estructura neuromuscular (nódulo sinusal) que genera los impulsos bioeléctricos que provocan las contracciones.

2.2.1. Ritmo Cardiaco

El ritmo cardiaco es la regularidad con la que tienen lugar los latidos del corazón. Dicho de otra forma, es el periodo armónico de latidos cardiacos que está formado por los sonidos de Korotkov [11]. El primer sonido tiene lugar en la fase de expulsión de la sangre del corazón (sístole), mientras que el segundo se genera durante el procedimiento de llenado del corazón (diástole).
Cuando tienen lugar dos sístoles y dos diástoles se produce un ciclo cardiaco. El ciclo cardiaco se define por tanto como el conjunto de sucesos cardiacos que tienen lugar desde que empieza un latido del corazón hasta el comienzo del siguiente latido.
En la figura 2.2 podemos ver una representación visual del concepto explicado anteriormente. El tiempo que transcurre para que se complete un ciclo se conoce como intervalo y su inversa es la frecuencia. Por esa razón, los corazones que tienen un

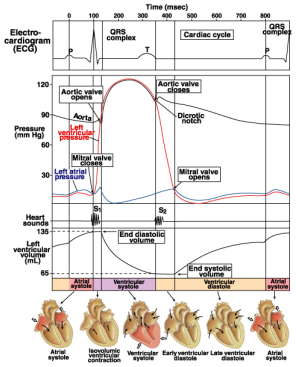


Figura 2.2: Diagrama que representa el ciclo cardiaco obtenido de [1]

ritmo de intervalos largos son los que presentan baja frecuencia (bradicardia) y los de cortos intervalos son los que presentan una alta frecuencia (taquicardia). Estos intervalos se miden en minutos, lo que conlleva a que la frecuencia se mida en Hertz (Hz=1/s). La frecuencia cardiaca se define como el número de contracciones ventriculares efectuadas por el corazón en un minuto. Se suele medir en latidos por minuto (lat.min-1) o pulsaciones por minuto (ppm). Normalmente la frecuencia normal en reposo oscila entre 50 y 100 latidos por minuto. Estas contracciones son la respuesta a las necesidades sanguíneas que precisa el organismo para poder satisfacer las funciones vitales.
Por lo tanto, controlando el ritmo cardiaco y la frecuencia se puede conocer si el paciente puede presentar algún tipo de alteración. Estas alteraciones tienen lugar cuando se producen cambios de la frecuencia o del ritmo cardiaco que no se justifican por razones fisiológicas. Pueden ser cambios que tienen lugar en la frecuencia cardiaca, ya sea porque se acelere (taquicardia) o porque disminuya (bradicardia). Pero en la mayoría de los casos, estas alteraciones suelen manifestarse cuando el paciente presenta un ritmo irregular. Es por todo esto que controlar el ritmo y la frecuencia resulta ser de vital importancia para nuestra salud.

2.3. Métodos para obtener el ritmo cardiaco a partir del análisis de video

Emplear métodos de análisis de video para detectar el ritmo cardiaco es un campo de estudio que comienza en 2002, con el uso de secuencias que surgen de cámaras térmicas [12]. A partir de este trabajo, surgen nuevos estudios que empiezan a utilizar cámaras convencionales. En 2007, se publica un método que utiliza secuencias en color generadas por una cámara convencional [17]. Sin embargo, es en el año 2012 cuando

empieza a producirse finalmente el crecimiento de este tipo de metodologías con la publicación de un estudio del Instituto Tecnológico de Massachusetts (MIT) [24]. Este trabajo expone que existe una variación de color en la cara ocasionada por el flujo sanguíneo. En consecuencia, se empiezan a realizar investigaciones relacionadas con la variación de color y posteriormente del movimiento. Finalmente, en el año 2013 se publica el primer trabajo que toma como referencia los movimientos producidos por la circulación sanguínea en la cabeza para calcular la actividad cardiaca a través del análisis de video [2].

2.3.1. Métodos basados en cambios de color

Este tipo de métodos calculan el valor del pulso a partir de las variaciones de color que se producen en la piel durante las diferentes fases del ciclo cardiaco. Estas variaciones de color son obtenidas de la región de la cara y posteriormente se les aplica un filtrado en la banda adecuada.
Existen muchas técnicas que se basan en los cambios de color para obtener el valor del pulso. Entre estas destacan:

- Comparación de frames en el espacio de color HSI (*Hue*, *Saturation*, *Intensity*) [7].
- Separación en componentes de color RGB de las que se realizará una normalización y un posterior Análisis de Componentes Independientes (ICA) o Análisis de Componentes Principales (PCA) [13, 8].
- Separación en componentes de color RGB a las que se les aplicará únicamente un procesado sin usar ICA ni PCA [20, 23] .
- Realizar la operación AND de la imagen binarizada en base a la clasificación del tono de piel mediante el uso de la escala cromática Fitzpatrick y la imagen en el espacio YUV [3].
- Descomposición piramidal y análisis de la componente R [4].

No obstante, estas metodologías dependen de las condiciones de iluminación ya que necesitan que la región de interés sea visible por la cámara.

2.3.2. Métodos basados en cambios de movimiento

El trabajo publicado en 2013 por Balakrishnan et al. [2] del Instituto Tecnológico de Massachusetts (MIT) fue el primer método que detecta el ritmo cardiaco a partir

de los movimientos de cabeza ocasionados por la presión de la sangre. El movimiento cíclico de la sangre desde el corazón hasta la cabeza pasando por la aorta abdominal y por la arteria carótida produce que la cabeza se mueva generando unos movimientos periódicos. Este algoritmo busca por tanto detectar el pulso mediante estos movimientos. La aproximación que plantean consiste en seguir puntos característicos en la cara de una persona, filtrar sus velocidades con una banda frecuencial temporal de interés y emplear PCA para encontrar una señal periódica generada por el pulso. Finalmente, se obtiene el pulso a partir de dicha señal y examinando su espectro frecuencial.

El primer paso del algoritmo consiste en localizar la región de interés (la cara) mediante el algoritmo de Viola-Jones [21]. Posteriormente, se obtienen los puntos de interés para la región escogida utilizando el algoritmo de *Good Features to Track* [15]. Después, se realiza un seguimiento de los puntos de interés en cada uno de los *frames* del video mediante el algoritmo KLT (*Kanade Lucas Tracking*) [10]. El resutado de este seguimiento serán las localizaciones serie-tiempo <x (t), y (t)> para cada punto. En esa investigación se ha decidido utilizar solamente la componente vertical de las trayectorias para realizar el análisis. Esto se debe a que los movimientos de cabeza relacionados con la actividad cardiaca son pequeños y están mezclados con otra variedad de movimientos involuntarios de cabeza. Por lo que las direcciones verticales resultan ser el mejor eje para medir el movimiento de cabeza vertical provocado por el pulso.

Una vez obtenida la componente vertical de cada punto a lo largo del tiempo (trayectoria), se procederá a realizar un filtrado de las mismas. El motivo por el que proponen realizar este filtrado se debe a que no todas las frecuencias de las trayectorias son necesarias o útiles para obtener la frecuencia cardiaca. El ritmo cardiaco de un adulto en reposo oscila entre [0.75, 2] Hz o [45,120] pulsaciones/minuto. No obstante, descubrieron que las frecuencias que se encuentran por debajo de 0.75Hz afectan negativamente en la representación del sistema que proponen. Mientras que los armónicos y las frecuencias que se encuentran por encima de los 2 Hz proporcionan una precisión útil y necesaria para detectar los picos. Por todo esto, decidieron que finalmente filtrarían las trayectorias con un filtro pasobanda comprendido entre [0.75, 5] Hz.

Finalmente, sobre las trayectorias filtradas se realizará un Análisis de Componentes Principales (PCA), que permitirá aislar el pulso. Para ello, se seleccionará la componente de movimiento que mejor se corresponda con el pulso basándose en su espectro. Es decir, se escogerá la componente que tenga una clara frecuencia principal. En la figura 2.3 se presenta un esquema que resume todo el proceso explicado anteriormen-

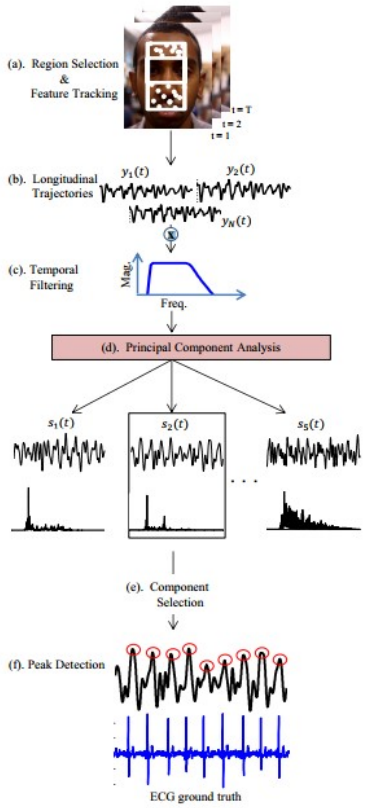


Figura 2.3: Esquema del cálculo del pulso cardiaco basado en cambios de movimiento. Figura extraída de [2].

te. En 2014, en la Universidad de Aalborg, Irani et al. [6] proponen modificaciones en el algoritmo propuesto por el MIT en 2013 [2]. Estas mejoras se implementan para las situaciones en las que el sujeto del video realiza movimientos de cabeza de forma voluntaria. Las transformaciones que introdujeron fueron añadir un filtro de suavizado después del filtrado de las trayectorias y en lugar de realizar la Transformada Discreta de Fourier (DFT) proponen usar la Transformada Discreta del Coseno (DCT).

Posteriormente, en el año 2015 Erik Velasco Salido utilizó el algoritmo publicado por el MIT en su Trabajo de Fin de Grado [19] y le añadió algunas extensiones. Estas transformaciones que realizó permitieron utilizar secuencias sin color proporcionadas por la Kinect para calcular el ritmo cardiaco. En este trabajo se emplea como algoritmo para detectar el ritmo cardiaco el propuesto por el MIT basado en cambios de movimiento [2]. Implementa dicho algoritmo como base para secuencias en color y

después desarrolla otro algoritmo para secuencias de profundidad y máscaras de segmentación. El algoritmo que desarrolla en imágenes de profundidad y segmentación efectúa la detección de la región de interés de forma manual. Es necesario mencionar también que para la obtención de las imágenes usadas en ese trabajo se utilizó como sensor la Kinect v1.

2.4. Kinect

La Kinect, conocida también como "Project Natal" es un mecanismo que fue creado como un simple controlador de juego y que fue lanzado en el año 2010 [18]. Este dispositvo está compuesto por una serie de cámaras y micrófonos. Es decir, no sólo es capaz de reconocer quién está delante de la cámara o que está diciendo, sino que también es capaz de detectar a qué distancia se encuentran los sujetos de dicho sistema. La Kinect está constituida por una cámara RGB, un sensor de profundidad, un micrófono multidireccional y un procesador personalizado. En la figura 2.4 se presenta la localización de estos elementos en la Kinect.

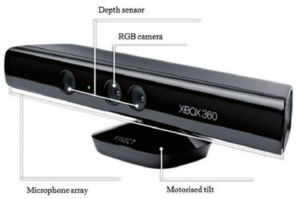


Figura 2.4: Componentes de la Kinect v1 [16]

La cámara RGB captura videos en color basándose en la teoría básica del color. Esta teoría sostiene que todos los colores se pueden formar mediante rojo (*Red*), verde (*Green*) y azul (*Blue*). Para formar dichas imágenes de color la cámara simula el comportamiento de nuestros ojos. Es decir, traduce las señales que se forman como resultado del reflejo de la luz al chocarse con los elementos que se encuentran en el escenario de grabación en colores.
Mientras que por otra parte, el sensor de profundidad es capaz de obtener la distancia a la que se encuentran los elementos que están presentes en el escenario de grabación. Este sensor está compuesto por un proyector de infrarrojos que está combinado con un sensor CMOS que es capaz de saber la distancia en 3D con independencia de las condiciones de iluminación de la zona bajo estudio.
En el año 2013 Microsoft anuncia la verisón 2.0 de Kinect. Son muchas las diferencias que presenta este nuevo sensor con su antecesor, las más relevantes se resumen en la tabla 2.1 .

Tabla 2.1: Diferencias entre Kinect v1 y v2

Funciones	Kinect v1	Kinect v2
video	640x480 30fps 1280x960 12fps	1920x1080 30fps
Profundidad	320x240 640x480	512x424
Rastreo del cuerpo	Detecta hasta 6 cuerpos y es capaz de rastrear 2 completamente. Identifica 20 articulaciones por cuerpo rastreado.	Detecta hasta 6 cuerpos y es capaz de rastrear todos completamente. Identifica 25 articulaciones por cuerpo rastreado.
Motor de Inclinación	Si	no
USB	2.0	3.0
Sistema Operativo	Sistema Operativo Windows 7 o superior	Windows 8.1 o superior y solo 64 bits
IR activo (visión nocturna)	No	si

Como se observa en la tabla 2.1, la Kinect v2 presenta una mayor resolución. Esto implica que se puede detectar con más detalle y precisión el entorno de estudio y obtener una mejor calidad de la imagen. Además, en la Kinect v2 el IR es independiente, pues en v1 se encontraba dentro del de profundidad. Esto permite tener una nueva fuente para manipular, la fuente infrarroja, que presentará una resolución de 512x424. Tanto la Kinect v1 como la Kinect v2 pueden proporcionarnos tres tipos de imágenes: en color (RGB), en profundidad (*depth*) e infrarrojas (IR). En la figura 2.5 se mostrará una representación del modelo v2.

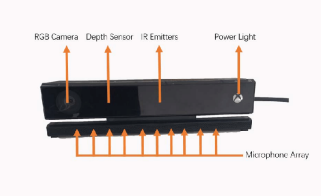


Figura 2.5: Componentes de la Kinect v2 extraído de [9]

2.4.1. Imágenes de Profundidad

La Kinect v2 es el dispositivo que hemos empleado para generar las imágenes del *dataset*. Este dispositivo como hemos explicado en el apartado anterior, tiene la capacidad de medir la distancia que hay entre la cámara y el entorno bajo estudio registrando así la distancia de los objetos que se encuentran en el contexto de grabación. Esto es posible gracias al sensor de profundidad que tiene integrado. Mediante la luz infrarroja, el sensor es capaz de generar una imagen de profundidad en la que

estarán capturados los objetos que se encuentran en el espacio.
En una imagen de profundidad el valor de intensidad de cada píxel indica la distancia a la que se encuentra esa parte de la imagen de la cámara. Estos píxeles presentarán un nivel de gris más o menos intenso (entre 0 y 255) en función de la distancia a la cámara. Por lo tanto, si el sujeto que se encuentra en la escena está cerca de la cámara en la imagen se verá más luminoso. Esto indica que en las imágenes de profundidad se representan las distancias más lejanas con niveles de intensidad cercanos a 0 (que es el negro) y las distancias cercanas con niveles de intensidad próximos a 255 (que es el blanco).
Así pues, al representar una de las imágenes de profundidad del *dataset* generado mediante la herramienta Matlab podemos observar lo representado en la figura 2.6.

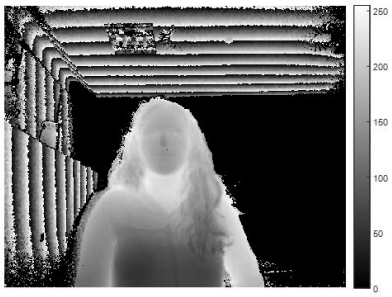


Figura 2.6: Imagen de profundidad obtenida del *dataset*

En esta imagen el sujeto bajo estudio se encuentra a 40cm de la cámara. Y como se ha explicado anteriormente, se puede comprobar que la cara y la ropa del sujeto al estar más cerca de la cámara se visualizan con niveles de gris entre 150 y 250. Mientras que el fondo de la sala en la que estaba el individuo, se presenta en negro por estar relativamente lejos.

2.5. Conclusiones

En consecuencia, por todo lo que hemos explicado anteriormente en toda la sección, se ha decidido emplear como método para calcular el ritmo cardiaco el análisis de movimiento en secuencias de video en profundidad obtenidas mediante el sensor Kinect v2. Pues en primer lugar, la versión v2 de la Kinect nos proporciona nuevas características y mejoras con respecto a la v1: una mejor calidad de la imagen obtenida y una mayor capacidad y precisión. En el último trabajo realizado [19] se empleaba la v1 para obtener las imágenes del *dataset*, por lo que hace que se abra

otro frente de estudio con el que obtener nuevos resultados para calcular el ritmo cardiaco.
Por otra parte, se ha decidido trabajar con imágenes de profundidad ya que permiten proteger la privacidad del individuo y ofrecer buenos resultados independientemente de las condiciones de iluminación. Además, durante la realización del estado del arte se ha corroborado que la mayoría de estudios encontrados para calcular el ritmo cardiaco a partir de análisis de video se basan en los cambios de color. Por lo que usar imágenes de profundidad para aplicar el algoritmo se converite en algo innovador.
En consecuencia, dado que estas imágenes solo nos proporcionan información espacial y no de color del espacio bajo estudio, se ha decidido utilizar un algoritmo basado en cambios de movimiento [2] para detectar el ritmo cardiaco.

Capítulo 3

Diseño y desarrollo

3.1. Introducción

Analizado el estado del arte, se ha decidido implementar lo más fielmente posible el algoritmo del MIT [2] basado en cambios de movimiento. En este capítulo se procederá a explicar el diseño y desarrollo de dicho algoritmo y de las mejoras implementadas para poder adaptarlo a imágenes de profundidad. Este algoritmo fue implementado en Matlab en el año 2015 por Erik Velasco Salido [19] como hemos explicado en el estado del arte. No obstante, aunque este algoritmo fue cumplimentado principalmente para aplicarlo a imágenes en color se implementaron mejoras para adaptarlo a máscaras de segmentación e imágenes de profundidad. Por lo tanto, para la implementación del algoritmo propuesto se ha decidido utilizar todo aquello considerado útil y válido para adaptarlo a la utilización de secuencias de profundidad que se haya encontrado en la publicación realizada por el MIT [2] y en el trabajo mencionado anteriormente [19].

3.2. Algoritmo Propuesto

3.2.1. Diseño

El algoritmo propuesto es por tanto una adaptación del algoritmo desarrollado por el MIT [2] al que se le han incorporado nuevas implementaciones para poder aplicarlo en imágenes de profundidad. Por lo tanto, se decidieron analizar tanto el sistema propuesto y seguido por el MIT [2] como el trabajo en el que se implementó el desarrollo del mismo [19] para posteriormente aplicar las mejoras que se hayan considerado oportunas para adaptar esta metodología a imágenes de profundidad. En la figura 3.1 se muestra el esquema seguido.

Figura 3.1: Diagrama algoritmo propuesto. La sección resaltada en naranja es en la que se han realizado nuevas implementaciones, mientras que las secciones en azul son adaptaciones del algoritmo propuesto en [2] .

Como hemos explicado en el estado del arte y resumiremos a continuación, el esquema consiste en lo siguiente. En primer lugar, se obtiene la región de interés en la que posteriormente se obtendrán los puntos de interés. Estos puntos de interés serán seguidos en cada uno de los *frames* del video y se obtendrán unas señales temporales que serán filtradas en una banda de frecuencias determinada. Sobre estas señales filtradas se aplicará PCA (Análisis de Componentes Principales) y se escogerá la señal pulso. Finalmente, se realizará la Transformada Discreta de Fourier de la señal pulso para obtener el valor de la frecuencia y en consecuencia el valor del ritmo cardiaco.
Como se ha dicho, se han tenido que modificar algunas partes del esquema planteado anteriormente para adaptarlo a la utilización de imágenes de profundidad. En primer lugar, se ha tenido que cambiar completamente la implementación de la primera etapa del esquema. En el algoritmo que proponemos se han aplicado otras metodologías para detectar la región de interés, ya que al Algoritmo de Viola-Jones [21] no es aplicable en imágenes de profundidad. Además, se ha mejorado esta sección, de forma que se puedan plantear dos opciones: seleccionar la región de interés de forma manual o que el algoritmo la seleccione de forma automática. El resto de etapas se han decidido mantener siguiendo la misma metodología.

3.2.2. Desarrollo

El algoritmo propuesto se ha desarrollado en Matlab. Para su implementación se ha realizado una función principal en la que se accederá al resto de funciones que componen el algoritmo. En esta función principal se muestran los videos presentes en el *dataset* y se ofrece la opción de poder introducir el video que se desea analizar. Hay dos tipos de videos: con todas las secuencias en profundidad y con el primer *frame* tanto en color como en profundidad y el resto de secuencias en profundidad. Por lo

tanto, si se escoge un video con secuencias solo en profundidad, se podrá calcular el ritmo cardiaco de forma manual o de forma automática. Mientras que si se selecciona un video que contiene un primer *frame* también en color, el pulso se obtiene de forma automática.

3.2.2.1. Algoritmo en modo manual

En este tipo de ejecución se aplicará el esquema planteado anteriormente en el diseño para el caso en el que la región de interés se selecciona de forma manual. Por lo tanto, se mostrará una ventana en pantalla para poder seleccionar la región de interés de forma manual. En la figura 3.2 se muestra una representación visual de lo explicado anteriormente.

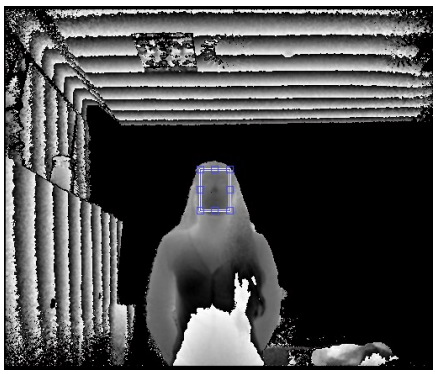


Figura 3.2: Selección de la región de interés de forma manual

Este método ofrece la flexibilidad de poder escoger la región de interés que se desee y obtener resultados específicos de dicha región. Por lo tanto, esta técnica ha sido un elemento clave a la hora de poder determinar cuál será la región óptima en cada una de las distancias que tenemos en el *dataset*.
Una vez detectada la región de interés, se procederá a realizar el cálculo de los puntos de interés. Se ha decidido emplear para ello la función *visión.CornerDetector* de Matlab que se basa en *Good Features to Track* [15] siguiendo lo expuesto en [2]. Se ha utilizado esta función como sistema de detección de puntos de interés por ser el único método con el que se pudieron obtener resultados aplicándolo en las imágenes de profundidad de nuestro *dataset*. Posteriormente, se planteará la opción de visualizar los puntos de interés encontrados o de seguir ejecutando el resto del algoritmo hasta obtener finalmente el valor del pulso cardiaco. En la figura 3.3 se muestra un ejemplo de lo que se obtendría por pantalla si se decide visualizar los puntos de interés para la región escogida.

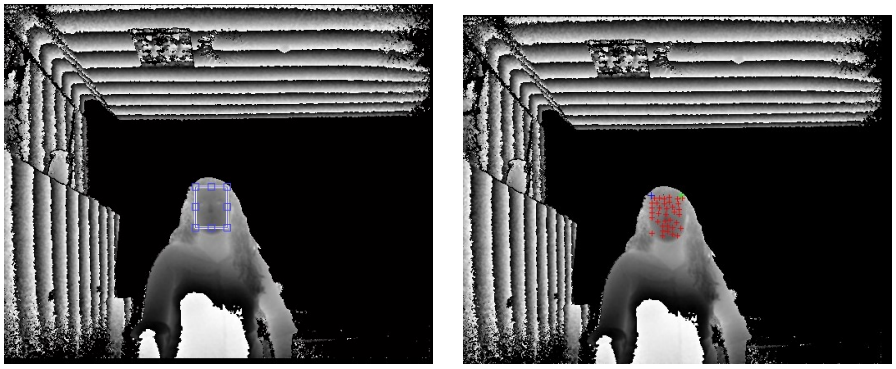


Figura 3.3: Representación en la imagen de la derecha de los puntos de interés encontrados para la región de interés seleccionada en la imagen de la izquierda.

Una vez obtenidos los puntos de interés, se realizará un seguimiento de los mismos aplicando el algoritmo de KLT [10] en cada uno de los *frames* del video. La función que se ha decidido utilizar para el tracking es *vision.PointTracker*, que realiza un seguimiento de los puntos de interés encontrados en el primer *frame* en el resto de secuencias. De esta forma, se obtendrá la variación de los puntos de interés a lo largo del tiempo. Sin embargo, se ha decidido trabajar únicamente con la componente y(t) atendiendo a lo mencionado en [2] y a lo explicado en el estado del arte. Posteriormente, se realizará un filtrado de cada una de las trayectorias obtenidas. Para ello, se ha desarrollado un filtro paso-banda Butterworth de orden 5 con frecuencias de corte comprendidas entre 0.75 Hz y 5Hz por lo esclarecido en el estado del arte. Aplicando el fitro implementado sobre las trayectorias se eliminarán las frecuencias no deseadas y se obtendrán las trayectorias filtradas. Sobre estas trayectorias filtradas se procederá a realizar un Análisis de Componentes Principales (PCA) seleccionando las cinco primeras señales generadas por este método. Se han decidido utilizar las cinco primeras señales obtenidas por PCA porque en el estudio publicado por el MIT [2] llegaron a la conclusión de que no es necesario considerar señales por encima de las cinco primeras para obtener buenos resultados.
Seguidamente, se ha realizado el cálculo de la Transformada Discreta de Fourier (DFT) para cada una de estas cinco señales relevantes con el objetivo de encontrar la señal que presente una mayor energía y clasificarla como puslo. Es decir, será la señal a partir de la cual se obtendrá el ritmo cardiaco. Para ello, se ha realizado el cálculo de la energía de cada uno de los espectros obtenidos para cada señal y la que presente el mayor valor de energía será la escogida. Posteriormente se buscará la frecuencia en la que tiene lugar ese valor de máxima energía, que será la frecuencia máxima. Por último, se ha obtenido el valor del pulso en pulsaciones por minuto multiplicando la frecuencia máxima por 60 segundos/minuto.

3.2.2.2. Algoritmo en modo automático aplicando Hough

En este método de ejecución, la función principal llamará a otra función en la que se seleccionará la región de interés automáticamente. Para que se pueda detectar la región facial de forma automática, se han desarrollado unas regiones estándar a partir del centro de la cara encontrada y de medidas antropométricas. El resto de etapas que se han realizado para concluir en el cálculo del pulso siguen el mismo procedimiento que en la opción manual. Por este motivo, la parte que más relevancia adquiere en este modo de ejecución es la que selecciona la región de interés de forma automática. En este modo de ejecución sólo podrán evaluarse las siguientes distancias: 60cm, 80cm, 160cm y 200cm y para cada una de ellas se ha generado una región de interés estándar. El motivo de la selección de estas distancias se explicará en detalle en el siguiente capítulo.

Para llevar a cabo la selección de la región de interés de forma automática se ha decidido utilizar la transformada de Hough [5]. La transformada de Hough es una técnica empleada para la detección de figuras en imágenes digitales. Mediante esta metodología se pueden encontrar todas aquellas figuras que puedan expresarse de forma matemática: líneas, circunferencias y elipses. Por lo tanto, este método nos permite buscar y encontrar circunferencias en nuestras imágenes de profundidad. La razón por la que se ha decidido buscar formas circulares y no elípticas es porque ha resultado ser la figura que mejor se asemeja con el contorno facial y que mejor ha encajado en el proceso de detección facial tras haber realizado varias pruebas. El desarrollo de esta técnica se ha aplicado en Matlab mediante la función *imfincircles*, que será explicada más en detalle a continuación.

Para poder detectar circunferencias en las imágenes de profundidad, primero ha sido necesario aplicar un detector de bordes. Estos detectores permiten entre otras cosas reconocer objetos, en este caso circunferencias. En el desarrollo de este algoritmo se ha decidido utilizar como detector de bordes Canny [14]. Se ha escogido este detector por ser de los mejores métodos de detección de contornos mediante máscaras de convolución y el uso de la primera derivada, además de ser de los más usados. Hay que mencionar también que se ha decidido aplicar en primer lugar una máscara a la imagen en la que se va a aplicar todo lo explicado anteriormente con el fin de aislar al sujeto bajo estudio del resto de la escena. El motivo por el que se ha decidido aplicar máscaras es para evitar falsas detecciones de cara y asegurarnos de que la mejor detección encontrada sea realmente la cara del sujeto de la escena. Estas máscaras serán diferentes en función de la distancia a la que se encuentre el sujeto y en todas ellas se asume que el sujeto se encuentra aproximadamente en el centro de la imagen. La función de Matlab que ha sido utilizada para aplicar el algoritmo de Canny

[14] sobre las imágenes se llama *edge*, y el resultado de aplicar esta función será una imagen contorno y se pasará como parámetro de entrada a la función *imfindcircles*. Esta última función será la que realice el cálculo de la Transformada de Hough aplicada a circunferencias, y devolverá tanto los valores de centro obtenidos para cada circunferencia como sus radios. Estos valores están ordenados en función de la fortaleza de las circunferencias. Es decir, las circunferencias que mejor encajen con los parámetros que se le han pasado a la función *imfindcircles* serán las primeras que se devuelvan. El otro parámetro que es necesario pasar a esta función es un vector que contiene los valores del radio mínimo y el radio máximo en el que se encuentran las circunferencias que se quieren buscar en la imagen contorno. Por todo lo explicado anteriormente, se ha decidido escoger como circunferencia elegida a la primera que devuelve este método. Pues dado que la cara de los sujetos tiene en la mayoría de los casos forma circular, debería ser la circunferencia con mayor fortaleza y la que mejor encaje con los valores de radio que han sido escogidos. Los valores de los radios inferior y superior escogidos han sido determinados en función de la distancia a la que se encuentra el sujeto. En la figura 3.4 se puede visualizar el resultado de aplicar todo lo explicado anteriormente sobre el primer *frame* de un video en el que se ha seleccionado este modo de ejecución.

Como se puede observar en la figura 3.4 en la imagen de la izquierda, la circunferencia representada es la que presenta mayor fortaleza y la que coincide justamente con la cara del sujeto. Esta imagen es el resultado de utilizar Canny [14] y posteriormente la Transformada de Hough [5] aplicada a circunferencias. Mientras que en la imagen de la derecha, se puede observar que la cruz que está representada en rojo en la imagen se corresponde con el centro de la circunferencia que se ha detectado como cara del sujeto y que se ha empleado para calcular la región de interés para este video. A continuación, en las figuras 3.5, 3.6, 3.7 y 3.8

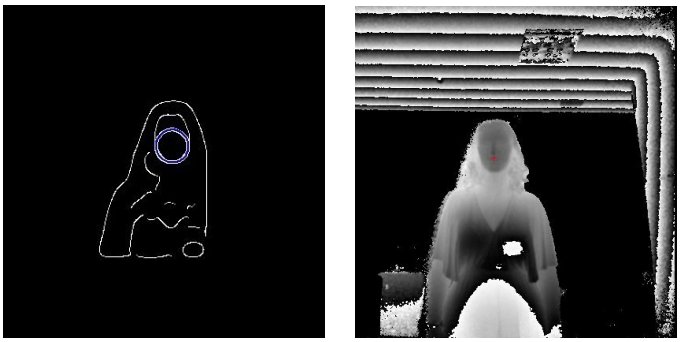


Figura 3.4: Detección de circunferencia con mayor fortaleza y representación de su centro

se muestran los resultados de aplicar esta misma función en videos que se encuentran a la misma distancia y a su vez a cada una de las distancias disponibles.

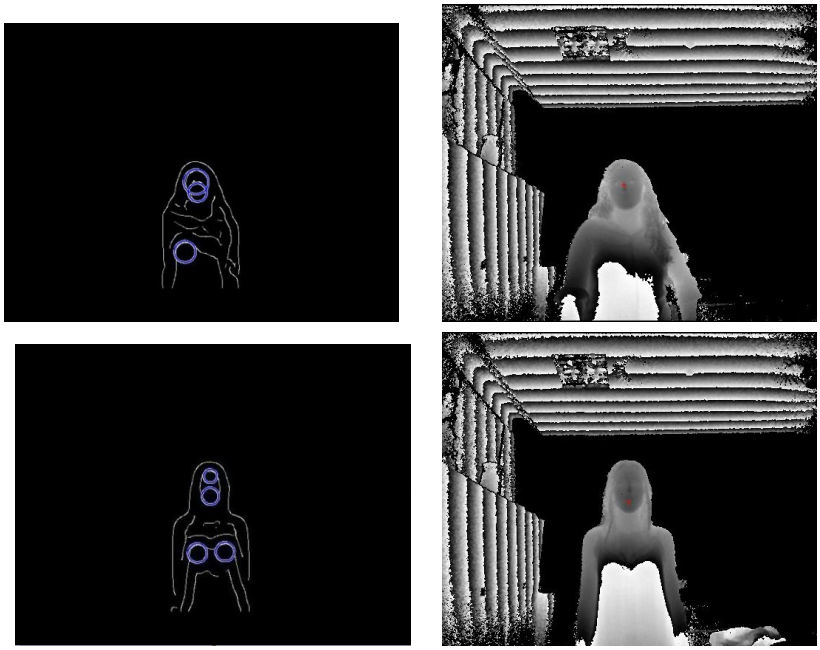


Figura 3.5: Resultados obtenidos en dos videos que se encuentran a 60cm

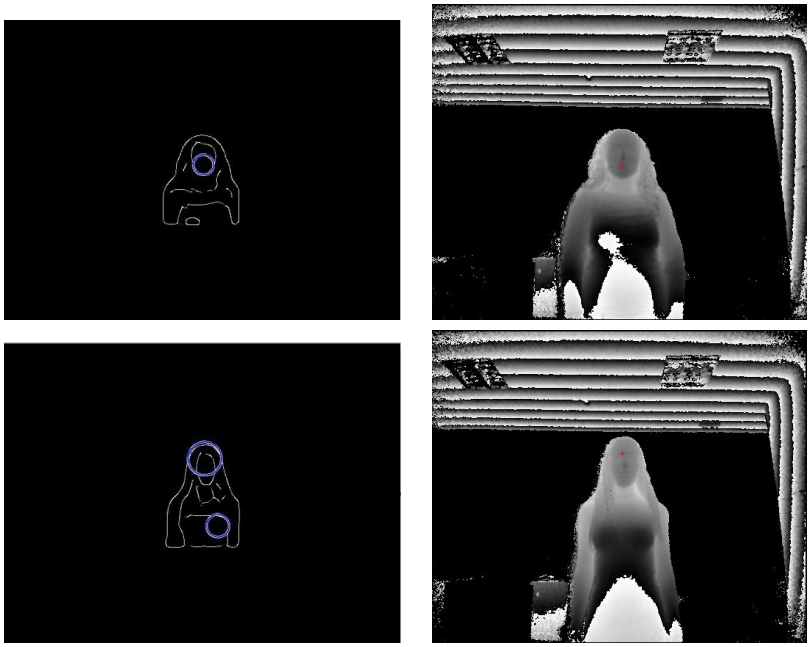


Figura 3.6: Resultados obtenidos en dos videos que se encuentran a 80cm

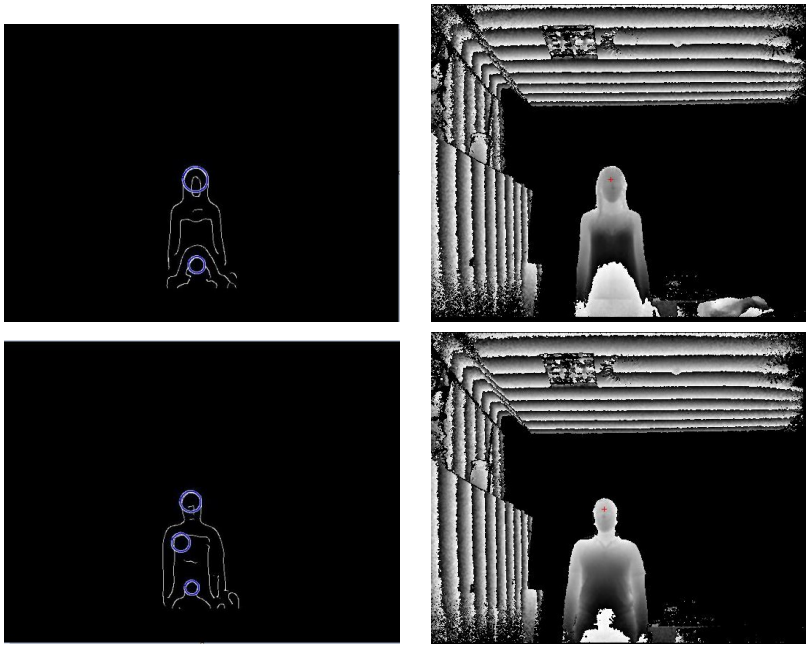


Figura 3.7: Resultados obtenidos en dos videos que se encuentran a 160cm

Figura 3.8: Resultados obtenidos en dos videos que se encuentran a 280cm

Analizando los resultados obtenidos en las figuras anteriores, se puede comprobar que aunque el algoritmo de Hough detecte otras circunferencias; la que mayor fortaleza tiene y la escogida es la que se corresponde con la región de la cara de los

individuos de la escena. Sin embargo, el centro de las circunferencias no se encuentra siempre en la misma posición para cada uno de los videos. Por lo que a la hora de calcular la región de interés estándar esto implica que es posible que en algunos casos la región escogida sobresalga un poco de la región facial a pesar de que los sujetos se encuentren a la misma distancia. Así pues, estas regiones estándar han sido determinadas en función de la distancia y de medidas antropométricas realizadas sobre varios videos que se encuentran a la misma distancia.

3.2.2.3. Algoritmo en modo automático aplicando Viola-Jones

Finalmente, se explicará el último modo de ejecución disponible. Este método es el que se ejecutará cuando se decida obtener el valor del pulso en secuencias que contengan el primer *frame* también en color. Por lo tanto, la explicación de esta metodología se basará fundamentalmente en la selección de la región de interés ya que el resto de etapas ya fueron explicadas en el método manual y se realizan siguiendo los mismos desarrollos. Dado que en este tipo de videos contenemos un *frame* en color se ha decidido utilizar el algoritmo de Viola-Jones [21] mediante la función de Matlab *vision.CascadeObjectDetector*. Así pues, la detección de la región facial se realizará mediante la función nombrada anteriormente y posteriormente se seleccionará dicha región en el primer *frame* de profundidad. Ha sido necesario cambiar la resolución de la imagen en color ya que presenta mayores dimensiones que la imagen en profundidad. En este método no ha sido necesario tomar en consideración la distancia a la que se encuentran los sujetos para encontrar la región de interés. Esto se debe a que el detector usado para reconocer la cara en las imágenes de color es capaz de encontrar las dimensiones apropiadas para la cara que ha detectado . En la figura 3.9

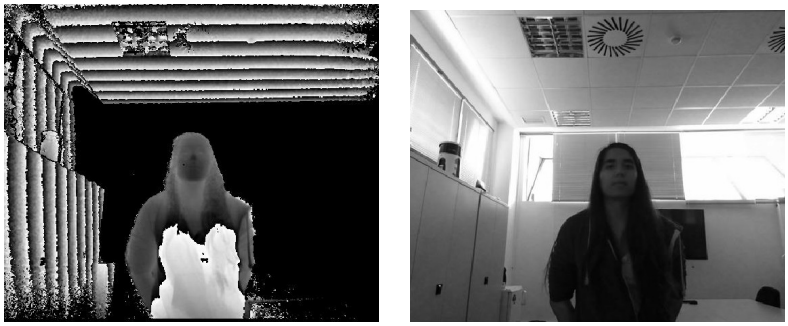


Figura 3.9: primer frame de un video obtenido tanto en profundidad (imagen de la izquierda) como en RGB (imagen de la derecha)

se mostrará un ejemplo de los primeros *frames* obtenidos en un video que contiene color y profundidad.

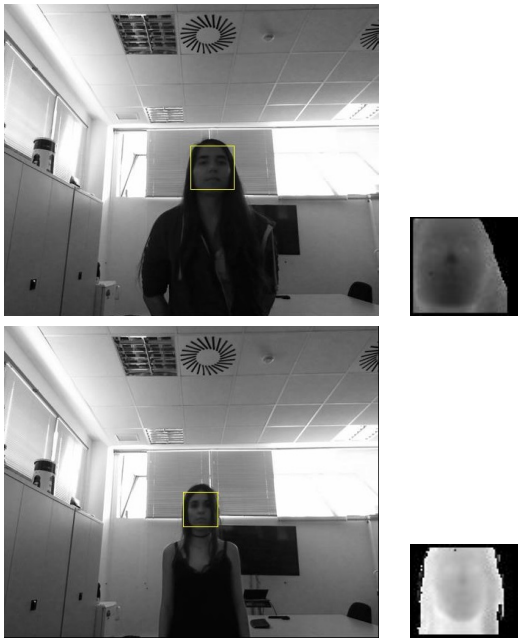


Figura 3.10: Detección de la región de interés a partir del algoritmo de Viola-Jones [21]

Así mismo, se podrá visualizar en la figura 3.10 la detección de la región de la cara en el *frame* de color y el resultado de aplicar dicha región en la secuencia de profundidad para dos videos diferentes.
Por lo tanto, como se puede observar en ambas figuras, el método empleado detecta correctamente la región de interés en los videos analizados y coincide en la imagen de profundidad. No obstante, en la región que se corresponde con la imagen de profundidad, se puede ver que tiene mayor dimensionalidad que en el caso del *frame* en color. Esto se ha intentado mejorar reduciendo esta región directamente en la imagen de profundidad, pero los resultado obtenidos eran peores que cogiendo la región obtenida al aplicar Viola-Jones en el *frame* en color.

3.3. Desarrollos no incluidos

3.3.1. Introducción

En esta parte se van a exponer algunos planteamientos que no se han llegado a implementar por su peor rendimiento o porque no han sido capaces de aplicarse y han tenido que ser descartados. Los desarrollos que no se han podido llevar a cabo han sido los siguientes:

- Escoger región de la frente y la región que se encuentra debajo de la nariz por ser posibles candidatas para obtener mejores resultados en el valor del pulso como se propone en [2].
- Detectar la cara de forma automática en las imágenes de profundidad empleando como método de detección de figuras Hough aplicada a elipses.
- Emplear otros métodos de detección de puntos de interés en las imágenes de profundidad (SIFT, SURF).
- reducir la dimensionalidad de la región de interés determinada mediante el algoritmo de Viola-Jones para el primer *frame* en color.

3.3.2. Desarrollo

3.3.2.1. Dividir la región de interés en dos subregiones

El uso de dos regiones de interés diferentes para obtener los resultados se ha descartado porque a pesar de que sea lo propuesto en [2] y sean consideradas como las regiones más estables, esta situación se planteaba para imágenes de color. Pero dado que en este trabajo estamos utilizando imágenes de profundidad, la selección de la región de interés más estable para obtener resultados puede diferir del escenario para secuencias en color. De hecho, se ha probado a utilizar estas regiones y los resultados obtenidos resultaron ser peores que escogiendo una única región de interés.

3.3.2.2. Detección de la región de interés mediante Hough aplicado a eplipses

Con respecto al segundo punto mencionado como posible desarrollo, no se llegó a implementar porque la región facial no era detectada en muchos casos como posible elipse. A pesar de haber realizado un realce de la imagen contorno obtenida con Canny con un filtro y de haber bajado el umbral en dicho detector de contornos se obtenían pocos resultados en los que la cara era detectada como elipse. Mientras que usando el detector de Hough aplicado a circunferencias siempre se detectaba la región facial como circunferencia.

3.3.2.3. Otros métodos para detectar puntos de interés

En lo que respecta a la detección de puntos de interés, se ha intentado aplicar otras técnicas de detección de los mismos para comprobar si los resultados obtenidos podrían ser mejores en función del método aplicado. Sin embargo, no ha sido posible

obtener puntos de interés para la región seleccionada como región de interés mediante estas metodologías. El único método que ha sido capaz de obtener resultados es el que se ha decidido utilizar en el algoritmo propuesto.

3.3.2.4. Modificar las dimensiones de la región de interés obtenida con Viola-Jones

Por úlitmo, como se ha explicado en la sección anterior se ha decidido no reducir o aumentar la dimensionalidad de la región de interés que ha sido seleccionada mediante Viola-Jones. Esto se debe a que al reducir dicha región y aplicarla en las imágenes de profundidad, se obtienen muchos menos puntos de interés y por lo tanto unos resultados peores. Mientras que si por el contrario se aumenta dicha región, se estarían considerando puntos de interés fuera de la región facial y de nuevo se consiguen peores resultados.

Capítulo 4

Evaluación

4.1. Introducción

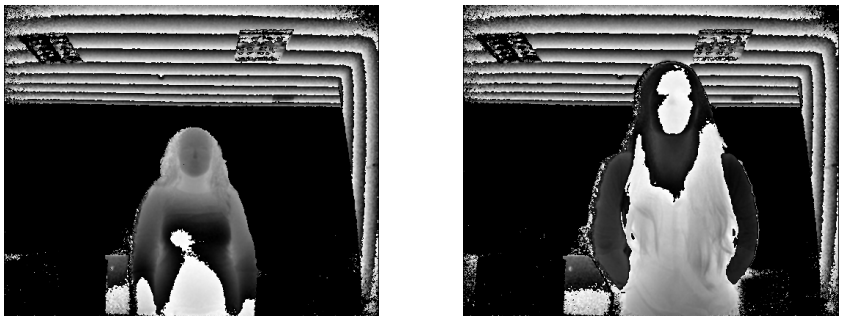
En este capítulo se explicará cómo se ha generado el dataset que se ha utilizado para la comprobación del algoritmo desarrollado. También se explicarán las pruebas que han sido necesarias realizar para llegar a obtener unos resultados con la mayor exactitud posible. Una vez obtenidos los resultados, se ha procedido a realizar un análisis de los mismos para poder extraer unas conclusiones que se detallarán más adelante.

4.2. Marco de evaluación

4.2.1. *Dataset*

El *dataset* que se ha utilizado está compuesto por dos conjuntos diferentes: los videos que solamente contienen imágenes de profundidad, y los que contienen el primer *frame* también en color. En ambos casos, los videos tienen una duración de 15 segundos.
Cada secuencia se ha grabado en dos modalidades diferentes: profundidad procedente de la Kinect v2 y color procedente de la Kinect v2 junto a su valor de *ground- truth* obtenido.
Como se ha explicado en el desarrollo del algoritmo, las secuencias en color sólo han sido obtenidas para el primer *frame* de forma sincronizada con el primer *frame* de profundidad en determinadas grabaciones del *dataset*.
Por otra parte, el valor de *ground-truth* de cada uno de los videos del *dataset* se ha obtenido realizando una media del pulso obtenido para cada sujeto durante la grabación de cada una de las secuencias. Este valor ha sido generado para poder

tener unos datos objetivos a partir de los cuáles verificar la precisión del algoritmo propuesto.
Para cada uno de los tres tipos de secuencias se han realizado grabaciones a diferentes distancias, siendo estas distancias: 40cm, 60cm, 80cm, 120cm, 160cm y 280cm. La razón por la que se escogieron estas distancias se debe al reseteo que se produce en la Kinect v2. Dicho sensor genera *frames* en profundidad que pueden llegar a no ser fiables, ya que en determinadas distancias la cámara se resetea y proporciona valores de intensidad en la escena que no se corresponden con lo que realmente se debería obtener. En las distancias en las que tiene lugar este reseteo se producen grandes cambios de contraste entre las figuras que se encuentran en la escena y el fondo. Por lo tanto, sabiendo esto se han realizado varias pruebas analizando las distancias en las que los *frames* obtenidos no eran fiables y eliminándolas del rango de evaluación en nuestro *dataset*. En la figura 4.1 se muestran un ejemplo de una imagen de profundidad con resultados no fiables y un ejemplo de una imagen de profundidad con resultados fiables.



(a) Imagen de profundidad con resultados fiables

(b) imagen de profundidad con resultados no fiables

Figura 4.1: Comparción entre dos imágenes de profunidad

Todas las grabaciones que componen este *dataset* se han realizado con 5 voluntarios diferentes y se han grabado en una única sesión por voluntario. Además, en cada una de dichas reproducciones el sujeto se encuentra posando frente a la cámara. Los medios empleados para poder llevar a cabo estas grabaciones han sido:

- Cámara Kinect v2 con una tasa de 30 *frames* por segundo y una resolución de 512x424 para las secuencias de profundidad y de 1920x1080 para las secuencias de color.
- Pulsímetro Fitbit Versa. Es un reloj de tipo pulsera que está sincronizado con la aplicación movil en la que se muestran en tiempo real las mismas pulsaciones que se están obteniendo en la pulsera en pulsaciones por minuto.

En la figura 4.2 se mostrará una representación de cada una de las distintas secuencias que componen el *dataset*.

Figura 4.2: Representación de las dos secuencias que componen el dataset

Para la grabación de las secuencias del *dataset* se ha utilizado el software en desarrollo por Julia Simón. Se ha colaborado conjuntamente para realizar la grabación de este *dataset*. Se decidió por tanto que los archivos de salida tuvieran la siguiente estructura:

- Depth_xxppm_xxcm, si el video contiene únicamente secuencias en profundidad. El primer parámetro será el valor *ground-truth* obtenido durante la grabación de la secuencia, y el segundo parámetro será la distancia del sujeto respecto a la Kinect v2.
- DepthandColor_xxppm_xxcm, si el video contiene también un primer *frame* en color sicnronizado con el primer *frame* de profundidad. El primer parámetro será también el valor de *ground-truth* obtenido durante la grabación del video, y el segundo parámetro será la distancia a la que se encuentra el inidividuo de la cámara.

El algoritmo propuesto utiliza el último parámetro para saber la distancia a la que se encuentra el sujeto en el video que está siendo analizado. En consecuencia, cuando se ejecute el modo automático se le aplica al video escogido la región de interés que le corresponde.

4.2.2. Métricas

Para poder realizar una evaluación de los resultados obtenidos a partir del algoritmo propuesto se han decidido emplear las siguientes métricas:

- Media del pulso medido. Se ha realizado una media del pulso medido durante la grabación de cada video. Se ha registrado el valor de este pulso medio en pulsaciones por minuto (ppm).

- Porcentaje de error. Como sabemos que el valor de pulso obtenido a partir del pulsímetro (*ground-truth*) es aproximadamente exacto, calcularemos el porcentaje de error que presenta el resultado que hemos obtenido con respecto al *ground-truth* mediante la siguiente fórmula 4.1 :

$$Error(\%) = |\frac{GroundTruth - pulsoObtenido}{GroundTruth}| \times 100 \tag{4.1}$$

- Porcentaje de error medio. Este valor se calculará como la suma de los porcentajes de error obtenidos entre el número de los porcentajes de error cuyo valor medio se quiere evaluar.

- Varianza. Para medir la dispersión de los datos mediante la siguiente fórmula 4.2 :

$$Varianza = \frac{\sum (x_i - \overline{x})}{n-1} \tag{4.2}$$

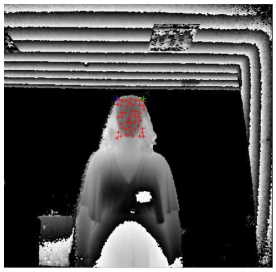
4.3. Pruebas y resultados

4.3.1. Pruebas y resultados del algoritmo propuesto en modo manual

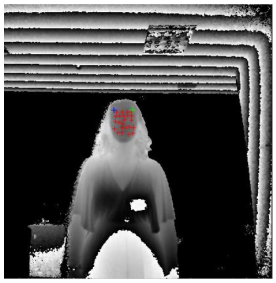
El objetivo de este primer grupo de pruebas ha consistido en determinar en qué región de interés se obtiene un resultado más cercano al valor del *ground-truth* obtenido para cada una de las grabaciones. El tamaño de las regiones estándar que se aplicarán en el agoritmo en modo automático se ha generado a partir de estas pruebas y resultados.

Por ello, en primer lugar se ha decidido analizar los resultados que se obtenían para un único video en el que se seleccionaban cuatro regiones de interés diferentes como se muestra en 4.3. El criterio para seleccionar estas regiones ha ido variando : comenzando con una región que abarca toda la región facial, hasta obtener una pequeña región en torno a la nariz. Este análisis de regiones se ha realizado para un tipo de video en cada una de las distancias. Se ha decidido escoger para mostrar los resultados de aplicar esta prueba un video en el que el individuo se encuentra a 80cm y su *ground-truth* es 71ppm.

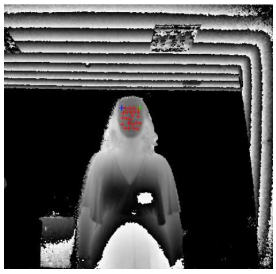
Como se puede observar en la figura 4.3, los resultados que más se aproximan al *ground-truth* son los que se han obtenido en las regiones 2, 3 y 4. En consecuencia, se ha decidido aplicar estas tres regiones sobre otras secuencias del *dataset* en las que el sujeto se encuentra a la misma distancia y analizar nuevos resultados que se recogerán en la tabla 4.1.



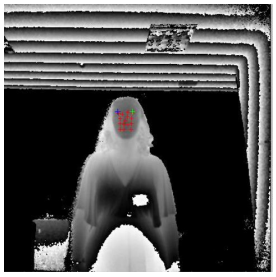
(a) Región 1 con valor de pulso: 46.0629ppm



(b) Región 2 con valor de pulso: 67.3228ppm



(c) Región 3 con valor de pulso: 70.866ppm



(c) Región 4 con valor de pulso: 70.866ppm

Figura 4.3: Representación de las cuatro regiones escogidas y los puntos de interés encontrados

Tabla 4.1: Tabla que recoge los resultados de aplicar las regiones 2, 3 y 4 en cinco secuencias de video que se encuentran a la misma distancia (80cm)

video	*Ground-truth (ppm)*	pulso en region 2 (ppm)	pulso en region 3 (ppm)	pulso en region 4 (ppm)	% Error en región 2	% Error en región 3	% Error en región 4
Depth_71ppm_80cm	71	67.3228	70.866	70.866	5,1791	0,1887	0,1887
Depth_75ppm_80cm	75	63.779	70.866	74.409	14,9613	5,512	0,788
Depth_84ppm_80cm	84	70.8661	74.409	85.039	15,6355	0,1141	1,2369
Depth_86ppm_80cm	86	109.842	95.669	63.779	27,7232	11,243	25,8383
Depth_87ppm_80cm	87	70.866	77.952	88.582	18,5448	10,4	1,8183
					16,4087	**5,4914**	**5,9740**

En primer lugar, es necesario analizar el *outlier* que se ha producido en los resultados para el video Depth_86ppm_80cm. Esto se debe a que en dicho video las secuencias de profundidad presentan un mayor contraste que en el resto de secuencias analizadas, por lo que los resultados obtenidos son peores como se puede observar. Se puede comprobar que la región que mejores resultados presenta es la 3, ya que es la que tiene un menor valor de error medio. En consecuencia, se ha decidido proceder a aplicar esta región sobre otros videos que se encuentran en el *dataset* y para otras distancias obteniéndose los resultados recogidos en 4.2.

Tabla 4.2: Resultados obtenidos aplicando la región de interés 3 en diferentes videos del dataset

video	Distancia (cm)	*Ground-Truth* (ppm)	pulso con región 3 (ppm)	% Error
Depth_65ppm_40cm	40	65	71.4586	9.936
Depth_82ppm_40cm	40	82	85.0394	3.706
Depth_94ppm_40cm	40	94	88.24	6.1276
Depth_82ppm_40cm	40	82	84.71	3.304
Depth_98ppm_60cm	60	98	95.29	2.7653
Depth_66ppm_60cm	60	66	52.94	19.7878
Depth_78ppm_60cm	60	78	77.65	0.4487
Depth_57ppm_60cm	60	57	56.47	1.7913
Depth_71ppm_80cm	80	71	63.7795	10.169
Depth_73ppm_80cm	80	73	74.4094	1.93
Depth_84ppm_80cm	80	84	81.4961	2,98
Depth_86ppm_80cm	80	86	88.5827	3.003
Depth_69ppm_160cm	160	69	60	13,0434
Depth_81ppm_160cm	160	81	84,4	4,1975
Depth_100ppm_160cm	160	100	98,82	1,18
Depth_73ppm_160cm	160	73	79,37	8,7260
Depth_80ppm_280cm	280	80	70.8661	11.417
Depth_82ppm_280cm	280	82	70.8661	13.577
Depth_75ppm_380cm	280	75	81,18	8,24
Depth_59ppm_280cm	280	59	56,47	4,2881

A partir de la tabla 4.2 se ha decidido generar la tabla 4.3, que recoge el porcentaje de error medio y la varianza para cada distancia. Se puede examinar en 4.3 que el

Tabla 4.3: Porcentaje de error medio de los resultados obtenidos en 4.2

Distancia (cm)	% Error medio	Varianza
40	5.7684	9.2748
60	6.1982	81.7787
80	4.5205	16.2122
160	6.7867	27.0154
280	9.3805	16.3304

porcentaje de error aumenta en función de la distancia, excepto para el caso en el que el sujeto se encuentra a 80cm. En esta distancia, se han obtenido los resultados más precisos ya que es la que presenta el menor porcentaje de error medio. Se debe tener en cuenta también que dado que los resultados se están obteniendo de forma manual, es

muy difícil seleccionar exactamente la misma región para todos los videos que hemos analizado y esto puede hacer que los resultados obtenidos sean menos concisos. Por este motivo, se ha decidido implementar el algoritmo de forma automática.

4.3.2. Pruebas y resultados del algoritmo propuesto en modo automático

Dado que la región con la que se obtuvieron mejores resultados de forma manual fue la número 3, en el modo automático se buscó generar unas regiones estándar para cada una de las distancias tomando como referencia la región 3. Sin embargo, se tuvo en cuenta que dichas regiones serán diferentes en función de la distancia a la que se encuentre el sujeto de la cámara. Cuanto más lejos se encuentre el individuo de la Kinect v2, menor será la región de interés.

En este tipo de pruebas se ha tenido que eliminar del proceso de evaluación los videos en los que el sujeto se encuentra a 40cm. Esto se debe a que no ha sido posible encontrar un método para detectar la región facial de forma automática. Después de haberse intentado aplicar tanto Hough aplicada a elipses como Hough aplicada a cirfunferencias, los resultados obtenidos no fueron satisfactorios. Se ha llegado a la conclusión de que dado que el sujeto se encuentra muy cerca de la cámara, la región facial no llega a identificarse como una forma circular o elíptica.

Se ha procedido por tanto a realizar pruebas para las siguientes distancias: 60cm, 80cm, 160cm y 280cm. Los resultados se han obtenido para tres videos diferentes en cada una de las distancias y se recogen en la tabla 4.4. En cuanto a los resultados

Tabla 4.4: Resultados obtenidos en el modo automático

video	Distancia (cm)	*Ground-truth* (ppm)	pulso (ppm)	% Error
Depth_98ppm_60cm	60	98	88,24	9,9591
Depth_78ppm_60cm	60	78	70,59	9,5
Depth_66ppm_60cm	60	66	56,47	14,4393
Depth_87ppm_80cm	80	87	77,65	10,7471
Depth_75ppm_80cm	80	75	67,06	10,5866
Depth_71ppm_80cm	80	71	67,06	5,5492
Depth_100ppm_160cm	160	100	98,82	1,18
Depth_73ppm_160cm	160	73	88,24	20,8767
Depth_69ppm_160cm	160	69	67,06	2,8115
Depth_75ppm_280cm	280	75	91,76	22,3466
Depth_67ppm_280cm	280	67	77,65	15,8955
Depth_59ppm_280cm	280	59	49,41	16,2542

obtenidos en 4.4, se debe tener en cuenta que para las mismas distancias se está usando la misma región de interés. Como se explicó en el desarrollo del algoritmo, los centros de la cara encontrados no siempre se encuentran en la misma posición para cada video, por lo que es posible que para algunas secuencias se estén considerando puntos que no son válidos. Esto hará por tanto que el valor de pulso obtenido difiera del *ground-truth*. Se ha decidido también generar otra tabla 4.5, en la que se recoja el porcentaje de error medio y la varianza en cada distancia. De nuevo se puede observar

Tabla 4.5: Porcentaje de error medio de los resultados obtenidos en 4.4

Distancia (cm)	% Error medio	Varianza
60	11.2994	7.4466
80	8.9609	8.7365
160	8.2894	123.3994
280	18.1654	13.1437

en 4.5, que los mejores resultados se han obtenido cuando el sujeto se encuentra a 80cm y a 160cm. En este caso, esto debe ser así porque las regiones estándar escogidas para esas distancias encajan mucho mejor con cada uno de los videos presentes en dichas distancias que las generadas para las otras dos distancias.

4.3.3. Pruebas y resultados del algoritmo propuesto en modo automático con color

El objetivo de estas pruebas es evaluar el valor del pulso obtenido por el algoritmo cuando se aplica en las secuencias que presentan el primer *frame* también en color y sincronizado con el de profundidad. Como ya se explicó en el desarrollo del algoritmo,aunque se hayan realizado las pruebas a diferentes distancias no es un parámetro a evaluar ya que la región que se ha seleccionado está adaptada a la distancia a la que se encuentran los sujetos en la escena. En la tabla 4.6 se pueden visualizar los resultados obtenidos.

Tabla 4.6: Resultados obtenidos en el modo automático con color

video	Distancia (cm)	Ground-Truth (ppm)	pulso (ppm)	% Error
DepthandColor_55ppm_60cm	60	55	67.06	21,9272
DepthandColor_57ppm_60cm	60	57	63.78	11,8947
DepthandColor_58ppm_60cm	60	58	49.61	14,4655
DepthandColor_61ppm_80cm	80	61	56,47	7,4262
DepthandColor_90ppm_80cm	80	90	120	33,3333
DepthandColor_53ppm_80cm	80	53	67.6	27.5471
DepthandColor_50ppm_160cm	160	50	45.88	8.24
DepthandColor_55ppm_160cm	160	55	56.47	2.672
DepthandColor_64ppm_160cm	160	64	45.88	28.3125
DepthandColor_59ppm_160cm	280	59	77.65	31.6101
DepthandColor_58ppm_160cm	280	58	73.52	26.7586
DepthandColor_51ppm_160cm	280	51	56.47	10.7254

Es necesario tener en cuenta que en todos estos videos la región de interés que se ha utilizado es algo más grande que la región 3 que se seleccionó como región óptima, por lo que los resultados obtenidos presentarán un error mayor con respecto al *ground-truth*.

4.3.4. Comparativa de los tres métodos

En esta sección se ha decidido comparar los resultados que se pueden obtener aplicando los tres métodos en los mismos videos y a diferentes distancias. Se mostrará en la tabla 4.7 los resultados globales obtenidos para cada método.

Tabla 4.7: Porcenta de error medio para cada cada método

Distancia (cm)	%Error medio en manual	% Error medio en automático	%Error medio en automático con color
60	7.6672	11.2994	16.0958
80	9.7121	13.4414	22.7688
160	8.7057	10.5395	13.0748
280	4.4038	21.3694	23.0313

En consecuencia, considerando los valores obtenidos en 4.7 se puede ver que los resultados son peores con los métodos automáticos que con el manual. Sin embargo, usar el método manual no es del todo preciso ya que es difícil seleccionar siempre la

misma región en diferentes videos o incluso en el mismo. En cuanto a los métodos automáticos se puede comprobar que funciona mejor el algoritmo basado en Hough que el que utiliza Viola-Jones. Además se puede analizar también que ambos métodos automáticos presentan un error notablemnte mayor en la última distancia. Por lo que podría decirse que a partir de los 160cm debería utilizarse el algoritmo en modo manual y no en modo automático.

4.4. Conclusión

Podemos concluir por tanto que los mejores resultados y más exactos son los que se han obtenido con el algoritmo en modo manual. Sin embargo, estos resultados se obtienen cuando se escoge como región de interés aquella que presenta unas características similares a la región de interés seleccionada como óptima en el capítulo. Por lo tanto, como es difícil seleccionar para cada una de las grabaciones siempre esta misma región se decidió implementar el algoritmo en modo automático. En base a los resultados obtenidos se ha podido concluir que el método automático funciona mejor utilizando Hough que Viola-Jones para detectar la región de interés.

Capítulo 5

Conclusiones y trabajo futuro

5.1. Conclusiones

El objetivo principal de este trabajo era implementar un algoritmo capaz de detectar el pulso en secuencias de profundidad utilizando métodos basados en cambios de movimiento. Después del estudio realizado en esta memoria, se puede concluir que el objetivo se ha cumplido.
También se puede decir que se ha cumplido el objetivo de implementar una metodología que sea capaz de detectar la región de interés en imágenes de profundidad. De hecho se han conseguido desarrollar dos técnicas: de forma manual y de forma automática. La primera técnica ha sido la base a partir de la cuál ha sido posible desarrollar métodos automáticos y con la que mejor resultados se han obtenido.
Se ha cumplido también con el objetivo de desarrollar un sistema que sea capaz de detectar la región de interés de forma automática. Se han implementado dos procedimientos en los que se obtenía la región de interés de diferentes formas: utilizando Hough aplicado a circunferencias o Viola-Jones en un primer *frame* en color. Se ha concluído que de estas dos metodologías la que mejor resultados ha generado es la primera. Además, se han adaptado ambos métodos a la distancia a la que se encuentra el sujeto de la cámara. Por lo tanto, se puede decir que se ha cumplimentado también el último objetivo. Este objetivo consistía en implementar al menos una técnica de detección facial automática que tenga en cuenta la distancia a la que se encuentra el individuo. Sin embargo, al tener que generar una región de interés óptima para cada distancia, se han obtenido peores resultados que con el modo manual. No obstante, siempre será mejor aplicar el método en forma automática antes que en manual, ya que la forma automática es capaz de encontrar la región de interés sin la interacción del usuario.

5.2. Trabajo futuro

A la vista de los resultados que se han obtenido en este trabajo se propone trabajar en:

- Profundizar en la detección de los puntos de interés en imágenes de profundidad. Intentar encontrar otras metodologías que detecten puntos de interés en imágenes de profundidad y que funcionen mejor que la que se ha implementado en este trabajo.
- Mejorar la detección de región automática basada en Hough. Intentar conseguir que el centro de las circunferencias encontradas se encuentre para todos los videos en el mismo sitio, para que luego al realizar las medidas antropométricas de la región estándar coincidan mejor en cada distancia.
- Ampliar el *dataset* con secuencias para un mismo individuo, para así poder evaluar si el error se mantiene constante o varía en los cambios de movimiento.
- Intentar buscar otra manera de conseguir que el *frame* en color y el *frame* en profundidad tengan las mismas dimensiones. De esta manera se conseguirá encontrar exactamente la región facial y no se tendrá que evaluar disminuir o aumentar su tamaño para obtener mejores resultados.
- Implementar todo este algoritmo en tiempo real. Esto se debería poder realizar en un lenguaje de alto nivel , como podría ser C++ haciendo uso de las librerías de OpenCV.

Bibliografía

[1] Diagrama de wiggers. 5

[2] Guha Balakrishnan, Fredo Durand, and John Guttag. Detecting pulse from head motions in video. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3430–3437. IEEE, 2013. 3, 6, 8, 12, 13, 14, 15, 16, 23

[3] Frédéric Bousefsaf, Choubeila Maaoui, and Alain Pruski. Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate. *Biomedical Signal Processing and Control*, 8(6):568–574, 2013. 3, 6

[4] L Carvalho, HG Virani, and S Kutty. Analysis of heart rate monitoring using a webcam. *International Journal of Advanced Research in Computer and Communication Engineering*, 3(05):6593–6595, 2014. 3, 6

[5] Juan Pablo Urrea Duque and Emmanuel Ospina. Implementación de la transformada de hough para la detección de líneas para un sistema de visión de bajo nivel. *Scientia et technica*, 1(24):79–84, 2004. 17, 18

[6] Ramin Irani, Kamal Nasrollahi, and Thomas B Moeslund. Improved pulse detection from head motions using dct. In *Computer Vision Theory and Applications (VISAPP), 2014 International Conference on*, volume 3, pages 118–124. IEEE, 2014. 8

[7] Alex Kleiner and Dan Rabinowitz. Video heart rate detection. 6

[8] Sungjun Kwon, Hyunseok Kim, and Kwang Suk Park. Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 2174–2177. IEEE, 2012. 3, 6

[9] Shenglan Liu, Muxin Sun, Wei Wang, and Feilong Wang. Feature fusion using extended jaccard graph and stochastic gradient descent for robot. *arXiv preprint arXiv:1703.08378*, 2017. 10

[10] Bruce D Lucas, Takeo Kanade, et al. An iterative image registration technique with an application to stereo vision. 1981. 7, 16

[11] Eugenia Urra Medina, Sandra Sandoval Barrientos, and Fabio Irribarren Navarro. El desafío y futuro de la simulación como estrategia de enseñanza en enfermería. *Investigación en Educación Médica*, 6(22):119–125, 2017. 4

[12] Ioannis Pavlidis and James Levine. Thermal image analysis for polygraph testing. *IEEE Engineering in Medicine and Biology Magazine*, 21(6):56–64, 2002. 3, 5

[13] T Pursche, J Krajewski, and Reinhard Moeller. Video-based heart rate measurement from human faces. In *Consumer Electronics (ICCE), 2012 IEEE International Conference on*, pages 544–545. IEEE, 2012. 6

[14] Jorge Valverde Rebaza. Detección de bordes mediante el algoritmo de canny. *Escuela Académico Profesional di Informática. Universidad Nacional de Trujillo*, 2007. 17, 18

[15] Jianbo Shi et al. Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, pages 593–600. IEEE, 1994. 7, 15

[16] Andrzej Skalski and Bartosz Machura. Metrological analysis of microsoft kinect in the context of object localization. *Metrology and Measurement Systems*, 22(4):469–478, 2015. 9

[17] Chihiro Takano and Yuji Ohta. Heart rate measurement based on a time-lapse image. *Medical Engineering and Physics*, 29(8):853–857, 2007. 3, 5

[18] Wouter van Teijlingen, Egon van den Broek, Reinier Könemann, and John GM Schavemaker. Towards sensing behavior using the kinect. In *Measuring Behavior 2012: 8th International Conference on Methods and Techniques in Behavioral Research*. Noldus Information Technology, 2012. 9

[19] Erik Velasco Salido. Detección de ritmo cardiaco mediante vídeo. B.S. thesis, 2015. 8, 11, 13

[20] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson. Remote plethysmographic imaging using ambient light. *Optics express*, 16(26):21434–21445, 2008. 3, 6

[21] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2001. 7, 14, 21, 22

[22] Neal Wadhwa, Hao-Yu Wu, Abe Davis, Michael Rubinstein, Eugene Shih, Gautham J Mysore, Justin G Chen, Oral Buyukozturk, John V Guttag, William T Freeman, et al. Eulerian video magnification and analysis. *Communications of the ACM*, 60(1):87–95, 2016. 3

[23] Lan Wei, Yonghong Tian, Yaowei Wang, Touradj Ebrahimi, and Tiejun Huang. Automatic webcam-based human heart rate measurements using laplacian eigenmap. In *Asian Conference on Computer Vision*, pages 281–292. Springer, 2012. 3, 6

[24] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. Eulerian video magnification for revealing subtle changes in the world. 2012. 1, 6